DOCUMENT RESUME

ED 187 752 TM 800 254

AUTHOR McKeegan, Hugh; Dennis, Cathy C.
TITLE Applied Performance Tests: How to Judge Them: How to Develop Them.
INSTITUTION Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
PUB DATE [75]
NOTE 37p.; For related document, see TM 800 253.
AVAILABLE FROM Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, Lindsay Bldg, 710 SW Second Ave., Portland, OR 97204 ($2.00)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; Information Sources; Measurement Objectives; *Minimum Competency Testing; *Performance Tests; Situational Tests; *Test Construction; *Test Selection
IDENTIFIERS *Applied Performance Testing

ABSTRACT

A good applied performance test (APT) concentrates on measurable objectives which are relevant for today's society in general, for the local community, and for the students. APTs should be valid, objective/reliable, straightforward in format, simple to understand, and able to maintain student interest. Scoring procedures should be standardized and objective, and administrative procedures should be efficient. The development of APTs generally involves eleven steps: (1) analyze content and objectives; (2) survey school and community; (3) decide on content; (4) analyze the task; (5) choose the testing situation; (6) develop directions; (7) develop scoring standards; (8) administer the test to a sample; (9) revise; (10) re-administer the test; and (11) determine its generalizability. Additional information about APTs may be obtained from teachers who have already developed tests, from the Clearinghouse on Applied Performance Testing (CAPT), from testing organizations such as Educational Testing Service, and from regional CAPT Materials Collection Centers. Selection of an APT from an outside source should be based upon the test's relevance to curriculum goals and objectives, as well as its technical qualities (such as reliability, validity, discrimination, format, relevance, scoring, and quality of instructions). (Six discussion questions and 24 bibliographic references are appended). (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE, NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED187752

APPLIED PERFORMANCE TESTS

HOW TO JUDGE THEM

HOW TO DEVELOP THEM

(Paper II)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Thorne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 8002254

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
Lindsay Bldg. - 710 S.W. Second Ave.
Portland, Oregon 97204

$2.00

Foreword

The Clearinghouse for Applied Performance Testing, among other functions, serves as a clearinghouse and information center for users or potential users of applied performance testing and assessment materials.

This paper was prepared for the Clearinghouse by Hugh F. McKeegan of the Department of Education at Bucknell University, Lewisburg, PA. He was assisted in this work by Dr. Cathy C. Dennis also of Bucknell University.

The principal purpose of this document is to provide information for elementary and secondary teachers who are interested in developing tests or in using applied performance tests developed by others. Additional references regarding applied performance testing may be secured by writing to CAPT at the address indicated on the cover page.

FEB 2 6 1980

Table of Contents

Foreword	p. 1
Introduction	p. 1
WHAT ARE THE CHARACTERISTICS OF GOOD APPLIED PERFORMANCE TESTS?	p. 2
HOW CAN I DEVELOP MY OWN APPLIED PERFORMANCE TESTS	p. 5
WHERE CAN I GET MORE INFORMATION ABOUT APTs AND THEIR DEVELOPMENT?	p. 18
HOW SHOULD APTs BE SELECTED?	p. 22
DISCUSSION GUIDE	p. 30
BIBLIOGRAPHY	p. 31

INTRODUCTION

This is the second of two papers prepared by the Clearinghouse for Applied Performance Testing to assist teachers in using applied performance tests in their classes. This document deals with the following questions:

- What are the characteristics of good applied performance tests?
- What steps should be followed in developing applied performance tests?
- What are the sources of information about APT?
- How should APTs secured from other sources be evaluated?

You may wish to read the paper entitled "Applied Performance Testing: What Is It? Why Use It?" before you begin this one. Comments and suggestions for improvement of either paper should be sent to the Clearinghouse for Applied Performance Testing, NWREL, 710 S.W. Second Avenue, Portland, Oregon 97204.

Please consider sending copies of any applied performance tests which you develop to the Clearinghouse so that they can be shared with other teachers. Your professional contribution will be acknowledged in the distribution of any such test.

WHAT ARE THE CHARACTERISTICS OF GOOD APPLIED PERFORMANCE TESTS?

The characteristics of an applied performance test can be discussed in terms of program, measurement, and instruction.

Program characteristics: By program characteristics are meant those attributes of a good test which are related to curricular outcomes or program goals. From the programmatic viewpoint, a good applied performance test is one which:

.Concentrates on those objectives of a program which are crucial in today's society.

.Concentrates on the kinds of objectives which can be measured effectively by student performance.

. Focuses on objectives which are considered critical by the community which your school serves.

.Emphasizes objectives which the students themselves see as important, relevant, and critical to success in life tasks.

Certainly not all applied performance tests will meet these criteria equally well, and a particular test should not be discarded because it only partially meets one or more of these criteria. A certain amount of subjectivity will enter into your judgment of a particular test, but a test which is weak in terms of all four criteria definitely should not be used as an APT in your program. If the test meets other needs, it may provide useful information for applications other than applied performance. Beyond this, you will have to weigh the strengths and weaknesses of

each test in terms of its programmatic use and the other criteria which follow.

Measurement characteristics: Characteristics and criteria related to measurement will be treated in more detail at a later point in this paper. Here four questions that are crucial to judging the quality of a test are suggested:

1. Is it valid? (Does the test measure what it purports to measure?)
2. Is it objective? (Can the test be presented in the same manner to all students? Have random factors and subjective judgments been eliminated or controlled?)
3. Is the format straightforward and simple to understand while maintaining student interest?
4. Are the scoring procedures standardized and objective?

There are many other measurement factors to be considered, but a test should rank high in terms of these four questions before it is given further consideration.

Administrative characteristics: Under this heading, you would consider factors that relate to the test's efficiency and ease of use by you as a teacher. Cost of use, both in terms of the purchase of instruments, simulations, and materials, and in terms of professional time in administration, scoring, and reporting, must be considered here. Even where school resources are such that cost of the test is not a factor,

you must examine the amount of time that the test will require from you, students, and other staff members. The extent to which the test will contribute to the up-dating and refining of curriculum objectives and instructional sequences should also be considered as a favorable factor.

An applied performance test that is strong programmatically and measurementwise should not be set aside because it presents some problem from the viewpoint of efficiency. This may be the very kind of test which you may wish to use to measure applied performance. A test that is effective can almost always be improved in its efficiency. An efficient test of dubious program or measurement quality should not be used because it will generate information that is incomplete, unreliable, or misleading and may mislead the student in dealing with similar situations in the future.

HOW CAN I DEVELOP MY OWN APPLIED PERFORMANCE TESTS?

After learning about APT, some teachers report that they have been using APT procedures in their classrooms under some other name without realizing it. You may be one of these teachers. Teachers in business and vocational programs or those of you responsible for work in consumer education, distributive education, art, and music may have more opportunities to develop and use applied performance tests than others.

The assumption here is that all teachers in all content areas, from kindergarten through grade twelve, can develop one or more applied performance tests on the basis of the material they are now teaching. What follows will be a response to you as an individual teacher interested in trying applied performance tests in your classes. Of course, if your school or school system is committed to developing an applied performance testing program, then a school-wide approach much more comprehensive than the procedures suggested here can be developed.

For the present, procedures that you as an individual teacher can carry out will be suggested. An eleven-step process will be presented. You need not follow all eleven steps with your first applied performance test, but eventually you will want to develop your tests in the complete form described here.

Table 1

Recommended Steps in Developing Applied Performance Tests

1. Analyze content and objectives.
2. Survey the school and community.
3. Decide on content.
4. Analyze the task.
5. Choose the test situation.
6. Develop directions.
7. Develop standards for scoring.
8. Administer the test to a sample.
9. Revise the test.
10. Re-administer to a sample.
11. Determine the generalizability of the test.

1. Content and Objectives Analysis.

You can begin with a careful review of the content and tests you are now using. What aspects of this content would be needed by a student who had to leave school at the end of eighth grade? Twelfth grade? Are there any items or groups of items in your current tests that could also be tested using applied performances?

List all those topics or skills that seem to be particularly amenable to either presentation in an applied performance context or testing in such a context. Now double check this list to make sure that all the topics you have listed (a) are related to life problems children must deal with now or face in the immediate future; (b) can be presented and/or require student response in an applied performance setting, either simulated or real. Do not delete any topics yet other than those which obviously do not meet these criteria.

Next, review any district course guides or curriculum guides to identify additional topics for applied performance testing. You may find some very good topics that are not currently included in your courses or are given only minor emphasis. Add these topics to your list.

Contact several of the sources of information about APT listed in a later part of this paper for more information about applied performance testing in general and for samples of APTs available at your grade level or in your subject area.

2. Informal School and Community Survey.

Check with teachers at grade levels above your own to determine what children can or cannot do. If you are a senior high teacher, talk to recent graduates, being careful to talk to both those who have gone on to further education and those who have not, about the problems they have encountered over the last six months. Talk to your principal; find out what complaints he has received about the school's instructional program. Speak to parents when they visit the school or attend parents' meetings. Ask them to tell you what, in their opinion, are the strengths and weaknesses of the school's program. Seek their cooperation in assisting you with either applied performance teaching or testing. In your interviews, do not overlook the students in your class. Talk to several of them about the knowledge and skills which they feel are important to success in life. Perhaps they may tell you about the problems older brothers or sisters may have encountered in getting a job, dealing with a government agency, finding a place to live, etc.

As a result of all this conversation, you will have several additional topics for possible applied performance testing to add to your list. Review the list again carefully. This time, narrow the list to five to ten topics that can be tested in the applied performance format and that other teachers, parents, school graduates, or your students agree are important. You should now have a set of topics or areas about which there is mutual school and at least some community concern. Now, rank order this final set of topics using such criteria as:

. relationship to life needs

. relationship to the goals of the school or program

. community interest

. your personal interest

. ease of preparing the test

Select two or three of the topics which rank high on these criteria for your first attempt at test construction.

3. Content Decision.

Carefully analyze each of the final two or three topics in terms of their content. Some behaviors associated with a given situation are extremely complex, and some decision must be made concerning which behavior should be tested and evaluated and which should not. For example, if you want to develop a performance test to measure the ability of a student to write a check, certain decisions must be made. Should attention focus upon the single behavior of actually writing a check or should allied behaviors, e.g., maintaining a current balance in a checkbook, ordering checks so that one has a current supply, etc., also be considered? Great care must be exercised in determining the content which is to be measured. Assume, for purposes of illustration, that you will be concerned with the rather simple behavior of writing a check to pay for some merchandise that has been purchased.

4. Analysis of the Task in the Situation.

After a decision has been made concerning the behavior

which is to be evaluated, the next step is the rigorous analysis of the task, i.e., the behavior, and the situation(s) in which the behavior will occur. The task must be analyzed carefully. For example, just what is involved in writing a check? First, the check itself must be filled in: a correct date must be written, often a number must be assigned to the check (some are pre-numbered), the name of the payee must be written in the proper space, the amount of the check must be printed numerically and written out in the proper places, and the person writing the check must provide a signature. Second, the check must be recorded properly in the person's records so that (a) the amount of the check can be deducted from the balance in the checking account; (b) the check can be identified when cashed so that the amount in the checking account can be verified; and (c) a reference concerning payment is available before the check is returned. Third, the individual should observe some basic procedures in writing checks: writing should be legible, the amount of money for which the check is drawn should be written carefully, and the purpose for which the check has been written should probably be noted on the check and in the records.

Likewise, the situation(s) in which checks are written should be rigorously analyzed. In what situations do people write checks? Sometimes checks are written in very quiet, comfortable surroundings, e.g., when people sit at home and draft checks to pay monthly bills. However, at other times, people write checks in different situations, e.g., in a bank or store, "payable to cash," for some pocket money, or in a crowded, busy

checkout lane of a grocery store, etc. Each situation in which the behavior is likely to occur should be studied carefully to determine its exact nature.

5. Choose the Situation in Which to Test the Behavior.

After the task has been analyzed and after a reasonable range of situations in which the behavior can be manifested has been identified and analyzed, a decision should be made concerning the situation(s) which will be used to measure the behavior. You must be careful to ensure that the situation(s) will remain rather stable. Assume that you desire to choose a situation which is rather complex: the writing of a check in a checkout lane of a supermarket.

The test can be one of two main types: direct observation and role-playing or simulation.[1]

One or both of the testing situations may be chosen to meet needs and objectives. If the objectives and content are clearly defined, and the task and situational analysis is complete, then the decision concerning the type(s) of test(s) and the actual development of the test is relatively straightforward.

---

[1]Viteles, M.S. Industrial Psychology. New York, NY: W. W. Norton Co., 1932.

Along with the development of the test, it will be necessary to determine the actual number of tests or situations that are needed. It seems reasonable that, in the supermarket example, only one supermarket would be necessary. But in a vocational class of thirty students, it may be necessary to have several engines to test for a specific engine problem. Again, a decision would have to be made regarding the number of instruments or tests needed.

Also, a decision will have to be made regarding the number of evaluators necessary for the test. In some situations, one evaluator will be sufficient, but in other situations, for economy of time, more than one evaluator may be necessary. It is suggested that in the latter case, advanced students who have already successfully demonstrated the tested competency be used as evaluators. An effective method for training peer evaluators is to role-play the part of the examinee while the competent student takes the part of the examiner.

6. Development of Directions for the Student.

After the situation has been chosen, appropriate direction must be developed for the student so that he/she fully understands what is expected. In the example discussed above, you might choose to use direct observation. Therefore, in the directions, you could inform the student that he/she will go to a designated supermarket, at a designated time when the supermarket will be quite crowded; will purchase a designated amount of specified items, e.g., two bars of soap, a quart of milk, a loaf of bread, a head of lettuce, etc.; will go to a busy checkout counter; will

wait in line; and will write a check, using the appropriate procedures taught previously, to pay for the purchase. These directions should be precise and clear so that the student understands completely what he/she is to do. It is suggested that the directions be in written form and followed exactly by the examiner. This does not mean that the student must read the directions, but the examiner should use them to make certain he has not forgotten portions of the directions. Also, if the directions are in written form, it is assumed that other examiners could give the test and maintain standard directions for all students. Where the test is to be administered to many individuals at separate times, the use of tape-recorded instructions should be considered. This will assure that each student receives a standardized set of instructions and may represent a considerable saving in teacher time.

In many schools, it might be impossible to conduct the test in a real supermarket because of restrictions in terms of time, money, administrative regulations, etc. In this situation, it might be possible to simulate the supermarket checkout counter in the school cafeteria or school store. It might be important to include in the simulation some of the distractions and time pressures which customers in a real supermarket checkout line encounter; writing a check without error within a certain time limit as a line of other student "customers" wait might meet this condition.

Remember that performance in the simulated situation gives us some greater assurance that students can perform in the real situation, but it does not guarantee such performance for all students. Where school regulations permit, it would be wise to select several students at random who have successfully completed the simulation and see if they can write an accept-

able check in the real supermarket. If some students pass the simulation but fail the "real test," then further practice in the simulation may be needed. It may also be possible that other problems which you will need to identify may be interfering with performance.

7. Development of Standards for Scoring.

One other task must be accomplished before the applied performance test can be used regularly: the standards which will be used to score the test must be developed. Each potential course of action which can be followed in the situation should be identified and described as precisely as possible. Then, effectiveness levels should be precisely indicated. For example, what courses of action will constitute success and what courses of action will constitute failure? Suppose that the student momentarily "freezes" in the real-life situation. How much time should be allowed before a decision to fail is made? Suppose the student writes the check but forgets to put a number on it. How should this be evaluated? Suppose the student writes the check and records it properly but does not note the purpose for which the check has been drawn on the stub. As can be seen, the various courses of action must be carefully analyzed so that effectiveness levels can be precisely determined. This can be an extremely difficult task. One must also determine whether the process and/or product will be evaluated and in what manner. After a scoring system is devised, a com-

plete written report should be prepared to assist others in using and scoring the test and to ensure uniformity in grading from one test administration to the next.

A brief statement of the criteria used to judge performance also should be prepared for examinees. Those who take the test should be informed in advance of what is expected of them and what is important in the test situation.

8. Administration to a Sample.

After steps one through seven have been completed, a random sample of students should be drawn from your class or program. The procedures devised as a result of the preceding steps should then be tested and followed in detail, to see if they actually are workable. There should be careful observation during the pilot testing in order to detect areas for improvement of the test. The observers in the pilot testing should record the score or scores for the students, and also maintain a record of how each student performed with respect to type of errors and unexpected responses. If you use both the scores and the records, your revision should be more productive.

When observers or different evaluators are brought into the classroom, you may get a special effect. Evidence is provided by many research studies, ranging from the Hawthorne Studies of the 1930's to the work of Rosenthal and Jacobson in 1968, that subtle but significant changes take place in the behavior of people who are being studied, whether by direct observation or other means. Behavior may change, for example, because

"observation satisifes certain needs of the observed, or because the expressed or presumed judgments of the observers have a suggestive power." Therefore, if you are using additional observers in the pilot or sample testing of your test, take precautions so that these observers do not add a new variable to your testing program.

9. Revision of Materials.

If the materials fail the test described in step eight, they must be revised. A careful study must be made of the behavior of the students in the sample to determine just where the problem(s) occur(s).

10. Readministration to a Sample.

After the revision described in step nine, step eight should be repeated. If necessary, steps nine and ten may have to be repeated again.

11. Generalizability of the Test.

After the material has been satisfactorily tested upon a sample and has been found to be suitable for use, a determination concerning the generalizability of the test must be made. Can students who have passed the test continue to write checks in supermarkets two months later or in other places which are busy, noisy, and confusing?

Steps eight, nine, and ten will consume a considerable amount of time, so you may wish to use your preliminary applied performance "test" in a practice mode with one or two classes over a marking period or semester before doing the final revisions. Your final goal should be to develop a test which can be administered, scored, and interpreted by someone else

without any intervention by you. Thus, older students, parents, other teachers, or teacher-aides can conduct the test so that your time may be devoted to other teaching responsibilites.

When APTs have been developed to this point, they can also be exchanged among teachers, further reducing time and effort in test development. Finally, do not overlook the possibility of having professional and business people or skilled craftsmen in your community participate in both test development and test administration.

WHERE CAN I GET MORE INFORMATION ABOUT APTs AND THEIR DEVELOPMENT?

This paper is just one of the references on applied performance testing available from the Clearinghouse for Applied Performance Testing! Among other documents available from CAPT are a bimonthly newsletter, references and bibliographies, status reports on various aspects of the field, a set of "Guidelines for the Evaluation of Applied Performance Test Materials and Procedures," and compilations of tests in adult literacy and other areas. In addition, other agencies and investigators have in many instances filed reports on their development of applied tests and related literature with CAPT. Thus the Clearinghouse has an ever increasing collection of materials dealing with aspects of applied performance testing of interest to teachers at the elementary and secondary levels.

Dissemination of information is only one function of CAPT. Among its other objectives are distributing applied performance testing materials and increasing the number of such tests available to schools by conducting development work. If you have been using your own applied performance tests in your classes, personnel at the Clearinghouse would be very interested in hearing from you. If your APT has proven successful for either group or individual use, you could make a professional contribution by making the test available to other teachers through CAPT. In turn, you would have access to those APTs in your areas of interest provided by other teachers. In submitting tests to the Clearinghouse, please enclose a brief

description of how the test was developed, its purposes, the grade-levels or programs in which it is used, and related information which would be useful to other teachers who may wish to use your test.

The Clearinghouse for Applied Performance Testing is a joint effort of the State Departments of Education of Oregon, Washington, Hawaii, and Pennsylvania, with support provided by the U. S. Office of Education. Thus, if you are a teacher in any one of the four states participating in the project, you will be able to secure certain applied performance materials from centers in your state. The addresses of these centers are listed at the end of this section.

There are many other organizations, particularly those that have participated in developing military and industrial training and vocational programs, who have a long standing interest in various types of applied performance assessment techniques. Prominent among these organizations are: Educational Testing Service, Princeton, New Jersey, and the Human Resources Research Organization, Alexandria, Virginia. You may also find that local vocational and technical schools and community colleges make use of applied performance procedures in their teaching and testing although they may not be labeled as applied performance tests. If you are reading this paper as part of a school in-service program, the director or leader of the program may be able to suggest local sites where applied performance tests are being used.

Should you be studying these materials on your own, and particularly if you are a teacher in some state other than Oregon, Washington, Hawaii

or Pennsylvania, then your best source of further information is to write directly to CAPT. Teachers in the member states may wish first to contact the Materials Collection Centers in their respective states for further information. Teachers anywhere who have developed usable applied performance tests should write to CAPT, Portland, Oregon 97204.

CAPT MATERIALS COLLECTION CENTERS

Hawaii

Hawaii State Library Branch
Social Science Section, 2nd Floor
478 South King Street
Honolulu, Hawaii 96813

Oregon

Clearinghouse for Applied Performance Testing (CAPT)
Northwest Regional Educational Laboratory
710 S. W. Second Avenue
Portland, Oregon 97204

The Resource Dissemination Center
Oregon Department of Education
942 Lancaster Drive, NE
Salem, Oregon 97310

Pennsylvania

College Library
Shippensburg State College
Shippensburg, Pennsylvania 17257

Research and Information Services for Education (RISE)
198 Allendale Rd.
King of Prussia, Pennsylvania 19406
(Attention - Ms. Emma Peterson)

Instructional Marerials Center
Annex 423
Temple University
Philadelphia, Pennsylvania 19122

Washington

Professional Curriculum Library
Office of the Superintendent of Public Instruction
Old Capitol Building
Olympia, Washington 98504

HOW SHOULD APTs BE SELECTED?

Evaluating APTs involves many of the same steps that you would use in assessing the quality of instructional materials and traditional tests for use in your classroom. In this section, we will concentrate on the evaluation of applied performance tests made available to you by other teachers or testing centers. Several steps proposed in the earlier section on test development represent a form of evaluation, sometimes called formative evaluation, which focuses on improvement and refinement of instructional materials and tests during developmental stages.

In examining instructional materials and tests from other sources, the nature of the evaluation differs from that used in developing your own materials and tests. You are evaluating the materials in terms of a "yes-no" decision as to whether to use them in your teaching. In this type of evaluation (summative evaluation), you will apply many of the same criteria that you apply to your own materials, but the purpose of the evaluation is primarily to make a decision regarding adoption rather than to improve or revise the instruments. Summative or decision-making evaluation can involve a myriad of questions, but several criteria seem to stand out as more critical than others in making an initial evaluation of an applied performance test. If the test "passes" in terms of these major concerns, then it can be examined further in terms of other important but perhaps secondary criteria.

Major categories of criteria in terms of initial evaluation should include:

Validity in terms of relatedness to important curriculum goals and objectives.

Technical qualities of the test.

Administrative efficiency.

If the test "passes" these major concerns, then it can be examined further in terms of important, but perhaps secondary criteria.

Relationship to Important Curriculum Goals and Objectives.

This involves an assessment of how well the test relates to outcomes of education which are particularly critical to the student in further education or in carrying out his responsibilities as a citizen and family member. While many of these application type outcomes, such as job-seeking, handling personal finances, and basic cognitive skills, are rather general to our culture, the priorities given to particular outcomes will vary according to the nature of the community and the socioeconomic background of the student. A test based on using the metropolitan area yellow pages to identify and locate private or government agencies that can assist with health or drug problems may be extremely useful to the urban student. For the student in the small town or rural area far from the city, such a test might be "fun" but it would not be very useful in helping him to secure assistance with such problems in his community. Thus, the test being considered must deal with important curriculum goals in a context which is realistic to the student now or in the very near future. (If a significant number of the graduates of the small town or rural high school move to the "big city" shortly after graduation, then some knowledge of how

to use the metropolitan telephone book could be quite critical!)

An APT that may be useful in testing for several important goals and objectives simultaneously should be given special consideration. Perhaps the test designer's primary objective was to determine whether the student could use a variety of resources to locate information about jobs. This, in turn, may involve reference to want-ads, telephone inquiries, visits to employment services, and job interviews. Reading skills, reference skills, verbal communication, both direct and via telephone, and ability to follow directions are all involved.

Such a test permits assessment not only of the survival skill of job seeking but also of the extent to which the student has mastered basic skills in reading and communication. It also provides diagnostic information useful to English teachers, remedial reading specialists, and guidance counselors. Tests that yield such a variety of data can be used both for the improvement of instruction in your classroom and for revision of curriculum at other grade levels or in other schools in your system.

Determining the relevance of an APT to curriculum goals important for your school is largely a logical and analytical task. It involves a careful review of the entire test, consideration of the extent to which it relates to critical goals and objectives, and an analysis in terms of its appropriateness and relevance to the community of which the school is a part. Good decisions in this respect involve a comprehensive knowledge of the curriculum and its objectives, the characteristics of the students who will use the test, and their home and community background.

Technical Qualities of the Test.

Technical qualities of the test refer to such factors as its validity, reliability, fairness, format, and method of scoring. A separate document treating these technical aspects of APTs in much greater detail than can be done in this introductory paper is available from the Clearinghouse for Applied Performance Testing. If you are reading this paper as part of an in-service program in your school, then the in-service leader or supervisor will probably have a copy of this reference. Otherwise, write to CAPT at the address indicated on the front cover and request a copy of "Guidelines for the Evaluation of Applied Performance Test Materials and Procedures."

A summary outline of the CAPT paper on test evaluation is provided on the following page. We will deal with only a few of the concepts from this outline in the remainder of this section.

Table 2

A Summary Outline of Guidelines for the Evaluation of Applied Performance Test Materials and Procedures[2]

1.0 Test Background

- 1.1 Purpose of the Test
- 1.2 Test Content
- 1.3 Task or Job Analysis
- 1.4 Pilot Testing and Validation

2.0 Characteristics of a Good Test

- 2.1 Validity
- 2.2 Reliability
- 2.3 Adequacy
- 2.4 Objectivity and Standardization
- 2.5 Comparability
- 2.6 Efficiency/Practicality
- 2.7 Balance
- 2.8 Difficulty/Discrimination
- 2.9 Fairness
- 2.10 Speededness
- 2.11 Format
- 2.12 Relevance

3.0 Test Administration and Reporting

- 3.1 Instructions to the Examiner
- 3.2 Instructions to the Examinees
- 3.3 Scoring

---

[2]Sumida, J. & Others, "Guidelines for the Evaluation of Applied Performance Test Materials and Procedures." 33 pages (mimeo) Portland Oregon: CAPT.

The outline on the previous page gives some indication of the complexity of a thorough technical evaluation of an APT or any other kind of test. Much of the information that you need cannot be gained by an inspection of the test alone. Questions such as the following need to be considered:

How reliable is the Test?

How good a predictor of future success is it? (predictive validity)

How closely does performance on this test correspond to performance on-the-job or in related test situations? (concurrent validity)

With what populations was the instrument field-tested and under what conditions?

How is this test scored?

To answer such questions, you may need to review manuals or other descriptive materials accompanying the test. Your technical evaluation of any test will be limited if these materials are not available. If a description of test development procedures, reliability, and validity data, etc., is not provided, you may still elect to use the test on a trial or pilot basis in your classroom. However, you will have to go through most of the steps previously suggested for developing new tests before you can use the instrument to make decisions about your students. Eventually test manuals or summaries should be available for all applied performance tests which are distributed to schools, but in the initial stages of development you may be given tests for which incomplete technical data is available. In this situation, it is critical that you rigorously analyze

the test in terms of its content validity. Does the content of the test, the knowledge and skills to be demonstrated by the student in the testing situation, adequately represent the range of behaviors required in a real performance situation? Content validity is easier to judge in those situations, such as in vocational education, where a thorough task analysis of the expected performance has been conducted. Where civic or social skills are being tested, simulation or the administration of the test in the real environment is a means of increasing the content validity of the test. Determination of content validity will demand your best judgment in assessing whether the test adequately represents the skills and outcomes expected in the real situation. If content validity is weak. then the test should not be used. Investment of time and effort in the use of such an instrument is a waste of resources. You are better advised to develop a new APT if content validity is low. If content validity is high, then one can consider revising the test to improve reliability, efficiency, scoring, etc. In technical evaluation of tests, it is also advisable to consult those teachers, counselors, or school psychologists who have specialized competence in tests and measurements and in the interpretation of technical manuals for tests. If you are experimenting with applied performance testing on your own, it may prove helpful to consult some of the references on tests and measurements cited in the bibliography.

Instructional Efficiency.

Factors of instructional efficiency should be considered only after the test seems acceptable in terms of its relation to important curriculum

goals and objectives and its technical qualities. An APT should not be accepted only because it is easy to administer or score. If the test is valid, reliable, and deals with important objectives and is also easy to use, then you have the best of the two worlds of effectiveness and efficiency. Many excellent APTs, however, may prove complex to administer and will require considerable time. Your decision in this instance must balance the effectiveness of the test in meeting critical objectives against the considerable investment of teacher and student time which may be involved in its use. Perhaps the judgmental question should be posed in the negative form, i.e., "Considering how good this test seems to be in measuring an essential survival skill, can I afford not to use it?"

Discussion Guide

1. Do you use standardized achievement or intelligence tests in your school? Who selected the particular tests you use and what is their purpose? What technical information is available regarding their quality?

2. How do you go about developing your own paper and pencil tests? In what ways are your test development procedures similar to those presented on pages 5 through 17? In what ways are they dissimilar?

3. Under what conditions would you use a paper and pencil test which had been developed by another teacher? What are the differences between using a test developed by another teacher and a textbook developed by another teacher?

4. Assume that you and several other faculty members all wished to experiment with APTS. In what ways could the group organize itself so as to be mutually supportive? What priorities would need to be established for test development and try-out? How would the exchange of tests and the use of tests from outside sources be arranged?

5. How could the cooperation of parents and community resource people be secured in the development of APTs? Would the use of APTs have a positive influence with parents and community members?

6. At this point, what additional information do you feel you need about applied performance testing? What local assistance do you need in order to begin developing APTs?

BIBLIOGRAPHY

1. CAPT Newsletter. Portland, Oregon: Clearinghouse for Applied Performance Testing, Vol. I, Vol. II, 1974-1975.

2. Sumida J., & Others. Guidelines for evaluation of applied performance test materials and procedures. Portland, Oregon: Clearinghouse for Applied Performance Testing, 1975, Mimeographed.

3. Viteles, M. S. Industrial Psychology. New York, New York: W. W. Norton Co., 1932.

Some References on Testing

. Bloom, B., Hastings, T., & Madaus, G. Handbook on Formative and Summative Evaluation. New York, New York: 1971.

. Boyd, J., & Shimberg, B. Directory of Achievement Tests for Occupational Education. Princeton: Educational Testing Service, April, 1971.

. Boyd, Jr., J. L., & Shimberg, B. Handbook of Performance Testing. Princeton, New Jersey: Educational Testing Service, 1971.

. Buros, O. K., editor. The Seventh Mental Measurements Yearbook, Volumes 1 & 2. Highland Park: The Gryphon Press, 1972.

. Cronbach, T. Essentials of Psychological Testing, 3rd Edition. New York, New York: Harper & Row, 1970.

. Gronlund, N. E. Preparing Criterion-Referenced Tests for Classroom Instruction. New York, New York: Macmillan Co., 1973.

. Osborn, W. C. Developing Performance Tests for Training Evaluation. Alexandria, Virginia: Human Resources Research Organization, February, 1973.

. Thorndike, R. L., & Hagen, E. Measurement and Evaluation in Psychology and Education. New York, New York: John Wiley & Sons, 1969.

Other References

1. Clearinghouse for Applied Performance Testing. References Related to Applied Performance Testing. Portland, Oregon: Northwest Regional Educational Laboratory, 1975, 64 pp.

2. Human Resources Research Organization. "Competency Based Education Brief." Alexandria, Virginia, 1974, 7 pp.

3. Jackson, R. Developing Criterion-Referenced Tests. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement and Evaluation (ETS), June, 1971, 21 pp.

4. Nafziger, D. H., Thompson, R. B., Hiscox, M. D., & Owen, T. B. Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments. Portland, Oregon: Northwest Regional Educational Laboratory, Assessment Projects, 1975, 131 pp.

5. National Assessment of Educational Progress. Math Fundamentals and Consumer Math. Denver, Colorado: Educational Commission of the States, 1975.

6. Nickse, R. S. Development of a Performance Assessment S,stem for the Central New York External High School Diploma Program, An Educational Alternative for Adults, A Progress Report in Applied Performance Testing. Syracuse, New York: Regional Learning Service of Central New York, Syracuse University Research Corporation, April, 1975, 30 pp.

7. Nickse, R. S. A Report on Adult Learners. A Profile of Fifty Adult Learners with Their Reactions to a Competency-Based Applied Performance Assessment System. Syracuse, New York: Regional Learning Service of Central New York, Syracuse University Research Corporation, November, 1975, 90 pp.

8. Northcutt, N. The Adult Performance Level Study. Austin, Texas: Division of Extension, The University of Texas at Austin.

9. Osborn, W. C. "Process Versus Product Measures in Performance Testing." Fort Knox, Kentucky: Human Resources Research Organization, October, 1974, 7 pp.

10. Panitz, Dr. Adolf, & Olivo, Dr. C. Thomas. Handbook for Developing and Administering Occupational Competency Testing. Washington, D. C.: U. S. Department of Health, Education and Welfare, Office of Education, National Center for Educational Research and Development, National Occupational Competency Testing Project, Volume 3, Phase I: Organizing - Planning - Pilot Testing, February 1, 1971.

Other References cont'd.

11. Pennsylvania Department of Education. Educational Quality Assessment in Pennsylvania: The First Six Years. Harrisburg, PA, 1973.

12. Swezey, R. W. & Pearlstein, R. B. Developing Criterion Referenced Tests. Reston, Virginia: Applied Science Associates, Inc., 1974.

13. United States Office of Education. Adult Functional Competency: A Summary. Washington, D.C., 1975.